

Towards Understanding User Tolerance to Network Latency in Zoomable Video Streaming*

Ngo Quang Minh Khiem

Guntur Ravindra

Wei Tsang Ooi

Department of Computer Science
National University of Singapore
Singapore 117417

{nqmkiem,ravindra,ooiwt}@comp.nus.edu.sg

ABSTRACT

We conducted a user study with 35 participants viewing 5 video clips to understand user tolerance to network latency when zooming and panning in zoomable video streams. With zooming or panning, unseen spatial regions in a frame are revealed and momentarily in an unknown state until data arrive from the server. To handle such unknown state, two common concealment schemes are used, namely Black scheme and Low-Res scheme. Black scheme renders the newly revealed region as black pixels, while Low-Res covers the unknown part with data from a low resolution video stream, which is additionally streamed by the server. In the context of these schemes, our study based on the simulation of delays shows that users are more tolerable to delay in Low-Res scheme. Up to 94% of participants can tolerate 1 second delay and 80% can tolerate up to a delay of 2 seconds in Low-Res scheme, while only 77% of participants can tolerate 1 second delay in Black scheme. The tolerable delay in zoomable video streaming is higher than thresholds found in some high interactive multimedia applications.

Categories and Subject Descriptors: H.5.1 [Multimedia Information Systems]: Evaluation H.4.3[Communications Applications]: Video

General Terms: Experimentation, Human Factors

Keywords: Zoomable Video, Delay Tolerance, Network Latency, Region-of-Interest Streaming, Concealment Scheme

1. INTRODUCTION

For the past few years, the proliferation of digital video cameras has widely enabled the capturing and publishing of High-Definition (HD) videos (1920x1080 pixels). Consumers, however, are increasingly watching videos on mobile devices that have limited screen size and display resolution. Due to the mismatch between captured and displayed resolutions, many of captured details are lost or cannot be clearly seen. This fact has led to the introduction of new interactions namely *zoom and pan* in video playback.

Zoomable video allows users to zoom and pan around a video to

watch a Region-of-Interest (RoI) at a higher resolution. Support for zooming and panning during video playback is easily achieved if videos are stored locally. We, however, are interested in supporting zooming and panning in the *streaming* context. When the client zooms in/out or moves a RoI, data corresponding to the new RoI co-ordinates needs to be fetched. Due to network latency, when users request for a RoI, they may experience a delay, between the instant the RoI changes and instant the data for the new RoI arrives.

In the server-client architecture, network latency is dominated by round-trip time while in peer-to-peer architecture, delay is due to various components which are influenced by protocol design. An attempt to characterize users' interactions with zoomable video was presented in [3]. One of the findings in that paper was that users actively and frequently interact (zoom/pan) with zoomable video. It was shown that 70% of the time the user interactions were spaced less than 1.6 seconds apart. In such cases, users may frequently experience a long waiting time before the new requested RoI arrives for viewing. Being kept waiting for every interaction may cause users to feel annoyed and lose interest in trying to use zooming or panning. Hence, the ultimate goal of zoomable video streaming system would be defeated.

Motivation. Past research on zoomable video has focused on encoding of zoomable video to support dynamic cropping of RoIs and understanding the influence of encoding parameters on zoomable video [7]. There is, however, no comprehensive study of users' tolerance to network latency in streaming of zoomable video. Such understanding is important since it can help in designing a system that is more responsive to frequent user interaction. Knowing tolerance levels to network delay helps us decide whether prefetching and caching are necessary for a streaming system of zoomable videos. For instance, if users can only tolerate a very low network delay (e.g., less than 1 second), prefetching or caching may be necessary to improve response time. In the context of peer-to-peer streaming of zoomable video, knowing maximum user tolerance to delay, we can build a better peer-to-peer streaming protocol for requesting or disseminating data of RoIs among the peers.

Approach. Being aware of the importance of understanding users' tolerance to network latency, we conducted a user study with 35 participants. The goal was to measure users' tolerable delay in two commonly used concealment schemes. The two concealment schemes attempt to minimize the wait time in response to a change in RoI. In the absence of the concealment scheme, the user interface waits for data belonging to the new RoI to arrive. During that interval, the video frame is frozen there by resulting in an unpleasant user experience. The concealment schemes attempt to quickly respond to a change in RoI. The first scheme moves the RoI to the new position as soon as user requests for a change. Newly revealed regions are rendered in 'black' as the data for these regions has not

*Area chair: Hari Sundaram

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.
Copyright 2011 ACM 978-1-4503-0616-4/11/11 ...\$10.00.

yet arrived. Once the data arrives, these regions are rendered with the correct pixel values. We refer to this scheme as *Black*. The second concealment scheme relies on the existence of a low resolution video stream which is always streamed by the server. In response to a change in RoI, the newly revealed regions are rendered with data from the low resolution stream. Once the requested data arrives, these new regions are displayed at the correct resolution. We refer to this scheme as *Low-Res*. Using these two schemes, we study user tolerance to delay in the context of zoomable video streaming.

The rest of our paper is organized as follows. Related work is discussed in Section 2. We describe our zoomable video interface, two concealment schemes in Section 3 and details of experiments in Section 4. Results of the user study are presented and discussed in Section 5. We conclude our paper in Section 6.

2. RELATED WORK

User tolerance to network latency and its effects has been well studied for many networked multimedia applications. ITU-T G.114 standards suggest that an end-to-end delay of at most 400 ms is recommended for general network planning [1].

In the context of streaming of progressive meshes, most users can tolerate up to 1 second of delay when the data rate is sufficiently high (100 Kbps) [4]. The reason for such high tolerable delay is due to the progressive nature of this application.

User tolerance to network latency and its effects in multi-player networked games has also been studied. Game genre differ in game nature and user interaction models, so does the user tolerance to delay. For Real Time Strategy (RTS) games which emphasize more on strategy, a study with the popular game Warcraft III found that users feel the smoothness for latencies between 0 ms and 500 ms. They start perceiving a degrade in game experience at some point between 500 ms and 800 ms [8]. For MMORPGs, a study with EverQuest 2 found that the game still runs smoothly with very high latency of 1250 ms [6]. For first person shooting, a study with the Unreal Tournament 2003 game has shown that a delay of 100 ms is noticeable while network latency of 200 ms is annoying [2].

3. ZOOMABLE VIDEO PLAYER

In this section we describe the zoomable video interface and two concealment schemes we used for the user study.

3.1 Zoomable Video Interface

We have previously implemented a web-based video player that allows users to zoom and pan around a High Definition (HD) video viewed in a small display. Figure 1 shows a screenshot of the zoomable video interface. On top is the main video display of small size 320x180, which simulates the scenario of watching a video under resource constraints (e.g., mobile devices with small screen size). A thumbnail window of size 160x90 is displayed in the bottom-left corner of the interface. The thumbnail view always shows a scaled-down version of the source video to provide users with the context of a RoI. The small rectangle on the thumbnail depicts a RoI. The region outside this small rectangle is made transparent so that the selected RoI more visible. The bottom-right corner of the interface shows control buttons for users to zoom in/out or move the RoI. Panning can also be done by dragging on the main display window, or clicking on the thumbnail display, while zooming is also possible with scroll wheel of a mouse.

Our zoomable interface can support six different zoom levels (0 to 5). The smallest zoom level 0 (by default) is equivalent to viewing the whole original video (1920x1080) at the display size 320x180 and hence, at the lowest level of detail. Watching a RoI

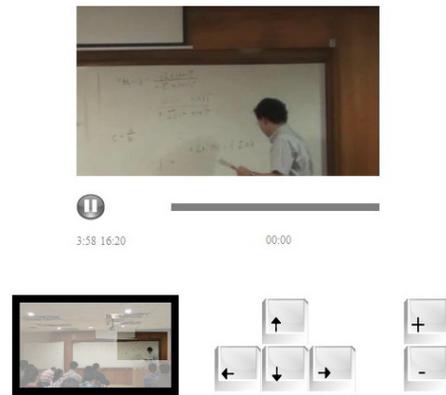


Figure 1: Snapshot of Zoomable Video Interface

at higher zoom level is equivalent to cropping a RoI at higher resolution and scaling it down to fit the display size of 320x180. For instance, at zoom level 5, a region of size 320x180 is cropped from original video and displayed. At level 4, a cropped region of size 640x360 from original video is scaled down to fit the main display.

3.2 Concealment Schemes

In this section, we explain in more detail the concealment schemes for which we measure the users' delay tolerance. When zooming or panning to change a RoI, a client sends the request and waits for the data of the new RoI to arrive. New RoI and the previous RoI may be overlapping. If so, part of the new RoI can be obtained from the previous RoI. Both *Black* and *Low-Res* concealment schemes maintain the continuity in viewing experience of users by displaying parts of RoI with data already available. In case the overlapping region comes from a RoI at different zoom level, upscaling or downscaling is necessary before rendering. Specifically, when users pan, the new RoI remains at the same resolution as the previous RoI. The overlapping part only needs to be re-located at its relative position within the new RoI on main display. In the case of zooming in, i.e., viewing a sub-region of current RoI at higher details, that sub-region of current RoI (at lower zoom level) is scaled up. With zooming out, i.e., viewing a larger RoI that contains the current RoI, the current RoI is scaled down and re-located at its correct position in the new RoI on the display.

The difference between Black and Low-Res concealment schemes lies in how they *conceal unavailable* parts of new RoI. Black scheme maintains the simplicity by rendering newly revealed part as region of black pixels while waiting for new data to come. Low-Res scheme, however, requires the additional transmission of thumbnail video, which is a very low resolution version of the original video (shown at the bottom left of our interface). Thumbnail video requires much less bandwidth and hence, should be readily available at client sides. Low-Res scheme covers any unfilled part of new RoI with pixels upsampled from the corresponding region of thumbnail video. Figure 2 illustrates how two schemes fill up the newly revealed region for each interaction (pan to move RoI to the right, zoom in and zoom out). Note that in the case of zooming in, there is no any newly revealed region. Regions seen as black in Black scheme appear blur in Low-Res scheme. It is due to the scaling up of pixels from thumbnail video.

In both schemes, users can still experience the smoothness when zooming or panning although they may see some blurring or blacking effects. By providing the responsiveness, we believe users' experience of delay is becoming less painful. We conducted a user

study described in the following sections to understand how users tolerate delays in Black and Low-Res concealment schemes.

4. USER STUDY

Video Clips. In this user study, we are using five different video clips captured by a HD camera. There are three videos of magic tricks (*Clock, Transfer, Dice*), one video of gymnastics performance (*Gym*) and one video of lecture (*Lecture*). In magic clips, the magician performs various tricks with the dices and the cards. For gymnastics video, a team of gymnasts perform on the floor of an indoor stadium. The lecture video captures a lesson in a university classroom where the lecturer uses the whiteboard to write notes. Participants can use our zoomable video interface to zoom or pan around in these videos to watch some region at higher resolutions. For example, the number on a dice or a card, the face or the movement of a gymnast, hand-writing of the lecturer on the whiteboard, etc. The camera is fixed and static in all clips. Videos are encoded with high quality to ensure the good viewing experience of RoIs at different zoom levels. Length of videos is about 3 to 5 minutes except lecture video of 16 minutes.

Experimental Setup. We conducted the user study on a PC with an Intel Core 2 Quad CPU 9550 at 2.83 GHz, 3 GB of RAM and running Windows XP. Firefox 3.6 browser was used to open the web-based player. All experiments were done in our research lab. The machine was connected to our central server. Test clips were pre-downloaded and stored on the local machine. This was to make sure the simulated delay was totally under our control and not affected by the actual network delay.

The simulation of effects of network latency in our system when users zoom or pan was implemented in the video player itself. Specifically, the video player delays the rendering of entire frames of newly requested RoI and shows only available parts of RoI instead. The entire frames are displayed after the delay expires.

Pilot Study. Before the actual user study, we conducted a pilot study with 8 users to find out the proper range of delay. We observed that most of these users rejected all test cases with delays over 4 seconds and none of them could distinguish among the delays below 1 second.

Experiment Parameters. Based on the pilot study, we used the delay range from 1 second to 5 seconds with the step size of 1 second. Each of these five delay values was *randomly* assigned with a video in the set of five clips. Different delay values were assigned with different videos. A delay and a video together formed a *configuration* and hence, we had 5 different configurations. We tested these configurations for the two concealment schemes. As such, a total of ten test cases were presented. Note that for different participants, the same delay value might not be associated with the same video. The fixed coupling between a delay and a video might result in the fairness issue among delay values due to the differences among the content of videos.

The same video content should not be watched multiple times in an experiment session. Being familiar with video content, participants may lose interest in watching or interacting with a video and hence, their evaluation of corresponding test cases are affected. As such, we used different videos for different delay values to alleviate this issue. Note that with current experiment setup, each video is viewed exactly twice in any experiment session.

Test cases were presented in a random order. There was no preference to any delay value or concealment scheme. We did not use the *method of limits* (i.e., start from some delay value (very low or very high) and keep increasing or decreasing the delay until the user's maximum tolerance level has been reached) [5]. For our study, this method is prone to the error of habituation since a partic-

ipant's tolerance level may have been adapted to the gradual change of delay and hence, is possibly improved beyond the actual level.

Experiments. A total of 22 male and 13 female participants, mostly from the university participated in this paid experiment. None of them had visual impairment or used this zoomable video interface before. Three participants, however, had previous experiences with zoomable video.

The participants were briefed about the user study and what to do in order to complete this experiment. They were first asked to read the instructions and watch a demo video on how to use this zoomable video interface. This was followed by a practice session to make sure they were familiar with different ways to zoom and pan before starting the actual experiment. In this practice session, there was no latency introduced in the interface, i.e. a participant would see the new RoI without any delay after performing zooming or panning actions. This was to *normalize* the differences among user expectations. Since not every participant had viewed zoomable video before, it was crucial that they should share the same expectation of a good streaming of zoomable video. The clip in this practice session was not used for the test cases.

Each participant was registered as a new session on the server. Users' interactions and answer choices were logged into our server database. For each session, five configurations (delay and video) were randomly generated to form 10 test cases. Ten test cases were presented to the participant in a random order. The delay value of each test case was not revealed. Participants were not even told about the presence of delay in these test cases. We, however, let the participants know that some blurring or blacking effects seen during zooming or panning were due to different implementations. For each test case, a participant is asked to watch and interact (zoom/pan) with a video. Each test case was followed by a question "*Do you find the responsiveness when zooming and panning acceptable?*" asking participants to evaluate the responsiveness of zooming or panning. The answer was given in the form of Yes/No Optional Buttons. The next test case is displayed if a participant selects and confirms his answers. Participants were not allowed to go back to previous test cases to change their answers.

5. RESULTS AND DISCUSSION

To understand how well users tolerate network latency when zooming and panning in two concealment schemes, we plot Figure 3. For every delay value in each scheme, we measure the percentage of participants who rated that delay value as *acceptable*. Note that it is allowed if a participant rejects some delay value while accepting longer delay values since user tolerance to delay may vary in different video content.

Results. The Figure 3 shows that for Black concealment scheme, only 77% of users can tolerate 1 second delay. There is a significant drop in acceptance percentage when the delay is 2 seconds (31%) and only less than 20% of participants can tolerate a delay up to 5 seconds for Black scheme. The delay tolerance levels of participants are significantly higher in Low-Res scheme. Up to 94% of participants can tolerate delay of 1 second and 80% can tolerate 2 seconds. Surprisingly, more than half of participants (51%) can tolerate up to a delay of 5 seconds.

Discussion. More users were tolerable to delays in the Low-Res scheme than Black scheme. This finding is as expected. In Low-Res scheme, when users change the RoI, they can see the newly revealed part of the RoI immediately, though at a low resolution. Even with 1 second or 2 seconds of latency, most users still found such delays acceptable. In Black scheme, the new part of RoI appears as black and hence, many users felt unpleasant when seeing nothing for new part of interest during waiting time. Low-Res

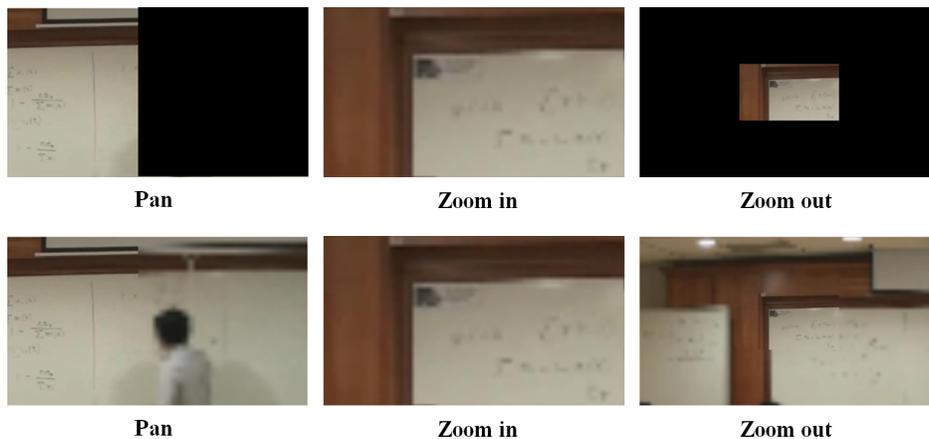


Figure 2: Concealment Schemes (Top row: Black scheme. Bottom row: Low-Res scheme).

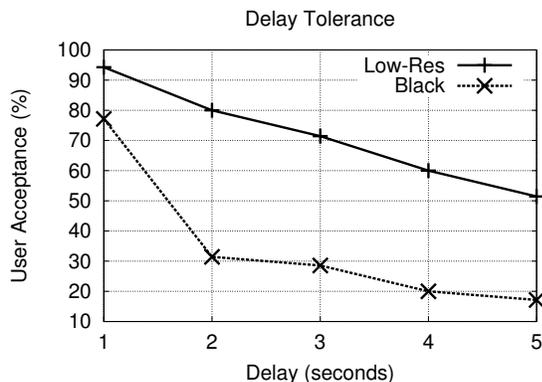


Figure 3: Delay Tolerance

scheme, while requiring the additional transmission of thumbnail video to viewers, has more users tolerable to delays.

The tolerable delay value in viewing zoomable video streams is higher than thresholds found in some high interactive multimedia applications, such as networked games. The reason for such a high latency is possibly because they can see the display of new RoI, in part or at low resolution, immediately. Both schemes attempt to "conceal" the presence of network latency by quickly responding to user interactions with zoomable videos. The users' tolerable network latency of 1 second is a good news for any peer-to-peer streaming systems of zoomable videos, since a requesting peer has more time to find other peers that have data for the entire or part of new RoI, or data are allowed to be forwarded through multiple hops to reach a requesting peer. Prefetching and caching in zoomable videos are necessary to quickly provide data of new RoIs in the presence of high network latency, since user tolerance starts degrading beyond 1 second.

6. CONCLUSION

This paper presented the findings of our user study on how much network latency users can tolerate in interactions with zoomable videos and how their tolerance levels degrade in the presence of network latency. We also show how the choice of concealment scheme helps improve delay tolerance levels of users. These find-

ings, we believe, can be incorporated into designing a system for streaming of zoomable video that provide both good Quality of Experience (QoE) and Quality of Service (QoS).

Acknowledgments

This research work is supported by National University of Singapore Academic Research Fund under the research grant WBS:R-252-000-368-112, and NExT Search Center, funded by the Singapore National Research Foundation & Interactive Digital Media R&D Program Office, MDA, under the research grant WBS:R-252-300-001-490.

7. REFERENCES

- [1] ITU-T Recommendation G.114 - one-way transmission time.
- [2] T. Beigbeder, R. Coughlan, C. Lusher, J. Plunkett, E. Agu, and M. Claypool. The effects of loss and latency on user performance in Unreal Tournament 2003. In *Proceedings of NetGames '04*, pages 144–151, Portland, OR, Aug. 2004.
- [3] A. Carlier, G. Ravindra, and W. T. Ooi. Towards characterizing users' interaction with zoomable video. In *Proc. of International Workshop on SAPMIA '10*, Florence, Italy, October 2010.
- [4] R. N. De Silva, W. Cheng, W. T. Ooi, and S. Zhao. Towards understanding user tolerance to network latency and data rate in remote viewing of progressive meshes. In *Proceedings of NOSSDAV '10*, pages 123–128, Amsterdam, The Netherlands, 2010.
- [5] G. Fechner. *Elements of Psychophysics: Volume 1*. Holt, Rinehart and Winston New York, 1966.
- [6] T. Fritsch, H. Ritter, and J. Schiller. The effect of latency and network limitations on MMORPGs: a field study of EverQuest2. In *Proceedings of NetGames '05*, pages 1–9, Hawthorne, NY, Oct. 2005.
- [7] N. Quang Minh Khiem, G. Ravindra, A. Carlier, and W. T. Ooi. Supporting zoomable video streams with dynamic region-of-interest cropping. In *Proc. of MMSYS '10*, Phoenix, Arizona, USA, 2010.
- [8] N. Sheldon, E. Girard, S. Borg, M. Claypool, and E. Agu. The effect of latency on user performance in Warcraft III. In *Proceedings of NetGames '03*, pages 3–14, Redwood City, CA, May 2003.